

January 13, 2025

Hon. Ona T. Wang  
United States Magistrate Judge  
Southern District of New York

Re: *The New York Times Company v. Microsoft Corporation, et al.*, 1:23-cv-11195; *Daily News, LP, et al. v. Microsoft Corporation, et al.*, 1:24-cv-3285

Dear Magistrate Judge Wang:

I write on behalf of Plaintiffs Daily News, LP, *et al.* (the “*Daily News* Plaintiffs”) and The New York Times Company (“The Times”) (collectively, the “News Plaintiffs”). We ask that you order OpenAI to preserve output log data on a going-forward basis for OpenAI’s generative AI products at issue, including ChatGPT, SearchGPT, APIs, Browse, and certain Custom “GPTs”, and to identify the nature and extent of output log data which it has destroyed at least since December 27, 2023, when The Times filed its initial Complaint.

### FACTUAL BACKGROUND

News Plaintiffs made it abundantly clear to OpenAI that output logs are relevant to the claims in these cases. The Times’s initial Complaint repeatedly asserts that the *output* of OpenAI’s products based on the GPT large language models directly infringes The Times’s copyrights, misappropriates its “hot news,” and/or dilutes its trademarks. *See, e.g., The New York Times Company v. Microsoft Corporation, et al.*, 1:23-cv-11195, Dkt. 1 at ¶¶ 98-101 (“[T]he current GPT-4 LLM will output near-verbatim copies of significant portions of Times Works when prompted to do so.”); *id.*, ¶¶ 102-07 (detailing example infringing outputs in section entitled “Unauthorized Public Display of Times Works in GPT Product Outputs”); *id.*, ¶¶ 108-23 (similarly providing examples of unauthorized retrieval and dissemination of current news); *id.*, ¶¶ 136-42 (referring to “‘Hallucinations Falsely Attributed to The Times’”). The *Daily News* Plaintiffs’ Complaint raises similar issues about the *output* of OpenAI’s GPT-based products with respect to their copyright, “hot news”, and trademark claims. *E.g., Daily News, LP, et al. v. Microsoft Corporation, et al.*, 1:24-cv-3285, Dkt. 1 at ¶¶ 96-138 and 170-76.

The Times also sent a letter to OpenAI stating that “OpenAI is obligated to preserve all documents related to ... output generated by the Models, including output that repeats or summarizes training material or contains hallucinations or misinformation ... [and] use of The Times’s content for retrieval augmented generation....” Declaration of Jennifer Maisel (“Maisel Decl.”), Ex. 1, pp. 1-2. On February 29, 2024, OpenAI responded to The Times, indicating that they were not going to preserve certain categories of output log data.<sup>1</sup> Maisel Decl., Ex. 2. In connection with this and other preservation issues OpenAI raised during discussions of the Parties’ Rule 26(f) report, The Times indicated that it “will comply with our preservation obligations and expect [OpenAI] to do the same.” Maisel Decl., Ex. 3. After these discussions, all Parties represented to the Court that the output of OpenAI’s models were relevant to the case in the Parties’

---

<sup>1</sup> No similar letter was sent to the *Daily News* Plaintiffs until November 15, 2024, as described below.

January 13, 2025  
Page 2

March 8, 2024, Rule 26(f) Report. *The New York Times Company v. Microsoft Corporation, et al.*, 1:23-cv-11195, Dkt. 72 at 2-4, 7, 10, 15, 17.

The News Plaintiffs issued several discovery requests related to OpenAI's output log data. For example, Request for Production No. 45 asks for "Query, session, and chat logs reflecting or analyzing user sessions related to the Publishers' Content within Defendants' Generative AI Products and Services, including, for each session, user queries paired with responses to those queries." Maisel Decl., Exs. 4-5. Similarly, Request for Inspection No. 6 asks for "Query, session, and chat logs related to [News Plaintiffs'] Content reflecting user sessions on Your Generative AI Products and Services ('Session Data'), including, for each session, user queries paired with responses to those queries." Maisel Decl., Exs. 6-7. OpenAI's responses and objections did not indicate that they were not preserving the requested information. Maisel Decl., Ex. 8-11.

It was not until OpenAI sent a letter to News Plaintiffs on November 15, 2024, that News Plaintiffs became aware that OpenAI was not preserving all relevant output log data that had long been requested and was the subject of on-going negotiations as it related to News Plaintiffs' requests for inspection. Maisel Decl., Ex. 12. News Plaintiffs sent a responsive letter requesting OpenAI to "(1) identify any output containing content from any of the News Plaintiffs that has been deleted; and (2) confirm in writing no later than December 17th, 2024 that you will hereafter preserve ChatGPT user conversations containing any of the News Plaintiffs' content." Maisel Decl., Ex. 13. News Plaintiffs asked OpenAI four separate times to meet and confer on the issue, and OpenAI has failed to agree or to confirm that it will preserve all relevant output data hereinafter.<sup>2</sup> Maisel Decl., Ex. 14.

## ARGUMENT

OpenAI has a duty to preserve information relevant to the asserted claims—which in these cases includes its output log data. "Once a party reasonably anticipates litigation, it must suspend its routine document retention/destruction policy and put in place a 'litigation hold' to ensure the preservation of relevant documents." *Zubulake v. UBS Warburg LLC*, 220 F.R.D. 212, 218 (S.D.N.Y. 2003). Accordingly, OpenAI had the duty to ensure the preservation of all ChatGPT user conversations, notwithstanding the retention period prescribed by OpenAI's data retention policy. *See Elavon, Inc. v. Ne. Advance Techs., Inc.*, No. 15-civ-7985, 2022 WL 1175039, at \*3 (S.D.N.Y. Apr. 20, 2022) (finding that Plaintiff did not satisfy its duty to preserve documents where it did not suspend its 90-day retention policy and "allowed its automatic deletion policy to run its course unabated, which lead to the destruction of its emails"); *see also Arista Recs. LLC v. Usenet.com, Inc.*, 608 F. Supp. 2d 409, 433 n.36 (S.D.N.Y. 2009) ("Courts have recognized the relevance of user log data and infringing digital files in copyright infringement actions, and they

---

<sup>2</sup> News Plaintiffs are pursuing a Rule 30(b)(6) custodial deposition with OpenAI on, *inter alia*, topics related to OpenAI's output logs. On a meet and confer on January 10, OpenAI indicated that the output log related topics were a "merits" issue and did not agree to put up any witness to speak on custodial issues relating to its output logs. Maisel Decl., ¶ 16.

January 13, 2025

Page 3

have ordered litigants to preserve and produce such data.”). OpenAI’s duty to preserve all ChatGPT user conversations arose no later than December 27, 2023, when The Times filed its initial Complaint. *See Arista Recs. LLC v. Usenet.com, Inc.*, 633 F. Supp. 2d 124, 139 (S.D.N.Y. 2009) (“There can be no dispute that the Defendants were under an obligation to preserve all documents and communications stored on their employees’ computers, at least as early as the start of this litigation.”).

OpenAI has not explained why it cannot preserve the data, or detailed the alleged burden of preserving the data. Data storage is cheap and abundant, even at the scale of data at issue here—and is a service provided, in fact, by OpenAI’s partner and co-defendant, Microsoft. However, because OpenAI has *not* been preserving all of its output log data as it is required to do, News Plaintiffs request that the Court order OpenAI to: (1) preserve its output log data, including output log data that contains News Plaintiffs’ content, hereinafter,<sup>3</sup> and (2) identify all output log data that has been destroyed, when it was destroyed, and whether it is retrievable.

January 13, 2025

Respectfully submitted,

/s/ Steven Lieberman

Steven Lieberman

Rothwell, Figg, Ernst & Manbeck, P.C.

/s/ Ian B. Crosby

Ian B. Crosby

Susman Godfrey L.L.P.

cc: All Counsel of Record (via ECF)

---

<sup>3</sup> The present motion concerns the preservation of output log data on a going-forward basis. News Plaintiffs are not now seeking sanctions or other remedies against OpenAI for its destruction of certain output log data since the inception of these lawsuits, but reserves the right to do so in the future.